The Effect of Deficiency at English on Female Immigrants' Wage in the UK

Miranda, Alfonso; Zhu, Yu

Published in:
Applied Economics Letters

DOI:
10.1080/13504851.2020.1753873

Publication date:
2020

Document Version
Peer reviewed version

The Effect of Deficiency at English on Female Immigrants' Wage in the UK: correcting for measurement error, endogenous treatment, and sample selection bias

Journal:	*Applied Economics*
Manuscript ID	AEL-2019-0846.R1
Journal Selection:	Applied Economics Letters incorporating Applied Financial Economics Letters
Date Submitted by the Author:	n/a
Complete List of Authors:	Miranda, Alfonso; Centro de Investigacion y Docencia Economicas, Economics Zhu, Yu; University of Dundee, Economic Studies
JEL Code:	J15 - Economics of Minorities and Races < J1 - Demographic Economics < J - Labor and Demographic Economics, J31 - Wage Level, Structure; Differentials by Skill, Occupation, etc. < J3 - Wages, Compensation, and Labor Costs < J - Labor and Demographic Economics, J61 - Geographic Labor Mobility|Immigrant Workers < J6 - Mobility, Unemployment, and Vacancies < J - Labor and Demographic Economics, C21 - Cross-Sectional Models|Spatial Models < C2 - Econometric Methods: Single Equation Models < C - Mathematical and Quantitative Methods
Keywords:	English as Additional Language (EAL), wage gap, endogenous treatment, sample selection

The Effect of Deficiency at English on Female Immigrants' Wage in the UK: correcting for measurement error, endogenous treatment, and sample selection bias

By Alfonso Miranda* and Yu Zhu†,‡

* Corresponding author. Economics Division & PANEL, Center for Research and Teaching in Economics (CIDE), sede Region Centro. Cicuito Tecnopolo 117, Aguascalientes 20313, Aguascalientes, Mexico. Telephone: +524499945150 ext. 5152. Email: alfonso.miranda@cide.edu. ORCID: 0000-0002-8176-0946.

† University of Dundee School of Business, Dundee, DD1 4HN, UK. Email: yuzhu@dundee.ac.uk. ORCID: 0000-0003-2805-4302.

‡ **Acknowledgements**: Both authors equally contributed to this paper. We are grateful to Jagjit Chadha, Antonio Di Paolo, Massimiliano Bratti, Amanda Gosling, Colin Green, Kai Liu, Stephen Machin, Anna Vignoles, Maria De Paola, Kjell Salvanes, Pravin K. Trivedi, João Santos Silva, Stephen Pudney, Gauthier Lanot, and Susan W. Parker for comments. **Disclosure statement**: The Authors declare that there is no conflict of interest. **Funding details**: The present research was done with no funding support. **Ethical approval**: We use secondary data from the the UK Household Longitudinal Survey (UKHLS), which is a longitudinal survey collected by ISER at Essex University, UK. The UKHLS has an Ethnicity Strand Advisory Committee, as well as a Scientific Advisory Committee, and complies with the 1964 Helsinki declaration and its later amendments.

Abstract

We investigate how English as Additional Language (EAL) affects the wage gap among foreign-born female immigrants in the UK. To deal with endogeneity and measurement error of EAL and self-selection into employment we implement a 3-step estimation (TSE) procedure suggested by Wooldridge (2002), which deliver consistent and asymptotically normal estimators. We find a large and statistically significant causal effect of EAL on the wage gap.

JEL: J15, J31, J61, C21

Keywords: English as Additional Language (EAL), wage gap, endogenous treatment, sample selection.

1 Introduction

Causal studies of host country language proficiency on immigrants' labor market outcomes are often plagued by the fact that language proficiency is measured with error and subject to individual choice (Dustmann 1994, Leslie and Lindley 2001, Chiswick and Miller 2014). An additional challenge with females is the need to account for strong selectivity into employment (Dustmann and Schmidt 2000).

Given the difficulties in implementing experimental or quasi-experimental identification strategies, we resort to an instrumental variable approach using data from the UK Household Longitudinal Survey. A novelty of the paper is the use of a three-step estimator, which allows for endogeneity and measurement error in a binary indicator of English as Additional Language (EAL) while correcting for self-selection into employment. Findings indicate that EAL carries a negative effect of between 30% and 33% for female foreign-born immigrants in the UK.

2 Data

We use the UK Household Longitudinal Survey (UKHLS) waves 1-6. As EAL is effectively time-invariant, we only use the last wave in which the respondent appears. UKHLS is a large UK longitudinal survey with information on ethnicity and country of birth of the immigrant and both parents, as well as one's first language. Our sample contains all female immigrants aged 19-59, who were born abroad to two foreign-born parents. Self-employed are dropped, as no earnings are available. After deleting missing values we have a sample of 5,381 females, of which 2,736 are salaried employees. Table 1 reports summary statistics. About 57% of females in work declare EAL, compared to 66% of the full sample.

3 Methodology

The model comprises six equations

$$EAL_i^* = \boldsymbol{x}_{i,EAL}\boldsymbol{\beta}_{EAL} + u_{i,EAL} \quad (1)$$

$$s_i^* = \boldsymbol{x}_{i,s}\boldsymbol{\beta}_s + \theta_s EAL_i + u_{i,s} \quad (2)$$

$$logw_i^* = \boldsymbol{x}_{i,logw}\boldsymbol{\beta}_{logw} + \theta_{logw} EAL_i + u_{i,logw}, \quad (3)$$

with,

$$EAL_i = 1(EAL_i^* > 0) \quad (4)$$

$$s_i = 1(s_i^* > 0). \quad (5)$$

$$logw_i = \begin{cases} logw_i^* & if \quad s_i = 1 \\ \text{missing otherwise} \end{cases} \quad (6)$$

Where $logw$ denotes log-wage, observable only if the employment selection indicator s_i takes on one; $i = 1, \ldots, N$. $EAL_i = 1$ if an individual is treated. The error terms follow a multivariate distribution with mean vector **0** and covariance **Σ**. Ordinary least squares (OLS) estimates are biased if $\mathrm{E}(u_{logw}|EAL) \neq 0$ or $\mathrm{E}(u_s|EAL) \neq 0$, in the presence of measurement error in EAL, or self-selection into employment. Previous work typically finds that the effect of EAL increases after correcting for endogeneity and measurement error (Yao and Van Ours 2019, Miranda and Zhu 2013).

In a model for a continuous response with a continuous endogenous explanatory variable and sample selection, Wooldridge (2002) recommends using a Heckman two-step approach to correct for the selection bias, while explicitly addressing endogeneity in the second step using a two-stage least squares (2SLS) estimator (Wooldridge 2002, p567). This is effectively a control function, which we term the Wooldridge's three-step estimator (Wooldridge 3SE). The method does not impose restrictions on the nature of the endogenous variable and can deliver a consistent estimator of (1)-(6) after dealing with the complication that the endogenous variable is binary. Following Wooldridge, we suggest fitting the second stage of the Heckman model by 2SLS instrumenting EAL with the fitted EAL probability from a first stage OLS regression to avoid the 'forbidden regression problem'. This gives rise to a modified 3SE. The first step fits the treatment EAL equation by OLS to get a predicted EAL probability. The second step is a reduced form that fits the selection equation by probit on all exogenous variables in the system—but not EAL—to obtain an inverse Mills ratio for correcting selection. The third step fits the wage equation by 2SLS using predicted EAL from the first step as instrument for EAL and controlling for the inverse Mills ratio from the second step and all exogenous variables in the system.

4 Identification

To secure identification we need two valid instruments, one for EAL and one for labor market participation. We postulate that the ***interaction*** between English-speaking country of origin and age-at-arrival to the UK is a valid instrument for EAL. Three identification assumptions are needed:

i. Learning a language as a native can only be archived within the early years of life.

ii. Controlling for age-at-arrival and country of birth, wages of immigrants in the UK who were born in non-English speaking countries is only affected through the language channel.

iii. After controlling for age-at-arrival and country of birth, any measurement error in EAL is uncorrelated with age-at-arrival*country-of-birth.

Under these conditions we have a valid instrument and Wooldridge 3SE delivers a consistent estimator that can be interpreted as a Local Treatment Effect on the Treated—where the control group are migrants who arrived from non-English speaking countries before age 10 and migrants from English speaking countries.

Assumption (i) is supported by a large body of evidence showing that children have the ability to learn any language to native fluency until they are about 10, when such 'plasticity' is lost (DeKeyser 2000).[1] Assumption (ii) ensures that our instrument is not correlated with the error term in the wage equation. It is important to underline that we do not require wages of immigrants in the UK to be unaffected by country of birth and/or age at arrival—in fact, we control for these. So, the assumption is plausible and other authors have use it before (Miranda & Zhu 2013, Yao & van Ours 2015, 2019). Finally, assumption (iii) ensures that our instrument corrects for measurement error in EAL.

The selection equation uses as instruments the female/male ratios of labor force participation and relative average educational attainment for population aged 25 or above in the country of origin. Both variables proxy for gender-based social norms of work. The main identification assumption is that social gender norms in the country of origin do not affect current wage in the UK, conditioning on

[1] We perform robustness checks using different ages as threshold and find similar results.

participation into the labor market, language skills (EAL), ethnicity, age-of-arrival, and religion. Yao and van Ours (2015) use the same variables to instrument employment selection in a study of language skills on immigrants' labour market outcomes in the Netherlands.

5 Results

Table 2 shows the effect of EAL on wages under different sets of controls. The raw wage gap is -0.192 log point. Accounting for differences in age, region of residence, wave, highest qualifications, dummy indicator for highest qualification obtained abroad, ethnicity, religion, and age-at-arrival (in columns 2 to 5) reduces the EAL coefficient to -0.134 log points; which remains statistically significant at 1%.

To investigate the relationship between EAL and labor market participation, Table 3 reports estimates from a bivariate probit model of EAL and selection into employment, allowing for non-zero correlation between the equations' disturbances. EAL enters the selection equation, but not vice versa, because the EAL status is a time-fixed characteristic set early in life—before women enter the labor market. In this specification our instrument for EAL strongly predicts EAL and in the selection equation higher female/male ratios of labor force participation rates in the country of birth are strong predictors of labor market participation. There is no evidence of weak instruments in either case. Table 3 also reports a statistically insignificant correlation coefficient between disturbances of -0.105.

Finally, Table 4 reports estimates from Wooldridge 3SE for the log-wage equation. To help the reader, we also report coefficients from OLS and 2SLS, reminding that 2SLS controls for endogeneity and measurement error of EAL but ignores potential sample selection into employment. The Wooldridge 3SE deals with all three problems. Robust standard errors for Wooldridge 3SE estimator are bootstrapped with 1000 repetitions.

Table 4 shows that correcting for measurement error and allowing for endogeneity of EAL more than doubles the size of the EAL effect, from -0.134 log points in OLS to -0.327 in 2SLS. Accounting for selection into employment leaves the effect effectively unchanged, going from -0.327 in 2SLS to -0.305 in Wooldridge 3SE. Hence, failure to account the endogeneity of EAL leads to an OLS estimate that is biased towards zero by more than half of the true effect—consistent with Yao and

van Ours (2019). However, ignoring self-selection into employment causes no major bias to the 2SLS estimator.

6 Conclusions

We address problems relating to measurement error and endogeneity of EAL along with potential self-selection into employment using Wooldridge's three-step estimator. Findings suggest that EAL has a causal effect of -30% to -33% on wages for female foreign-born immigrants, which is significant at 1%. This is more than twice as much as the OLS estimate of -13%. Our results also suggest that EAL has no bearing on women's labor market participation, conditional on other controls and once the endogeneity of EAL is explicitly controlled for.

Appendix

Supplementary material for this paper can be found online.

References

Chiswick, B.R. and Miller, P.W. (2014) International Migration and the Economics of Language. In: *Handbook on the Economics of International Migration*. Elsevier.

DeKeyser, R.M. (2000) The Robustness of Critical Period Effects in Second Language Acquisition, *Studies in Second Language Acquisition* 22, 499-533.

Dustmann, C. (1994) Speaking fluency, writing Fluency and earnings of migrants, *Journal of Population Economics* **7**, 133-56.

Dustmann, C. and Schmidt, C. (2000) The wage performance of immigrant women: Full-time jobs, part-time jobs, and the role of selection, *IZA Discussion Paper* No. 233.

Leslie, D. and Lindley, J. (2001) The impact of language ability on employment and earnings of Britain's ethnic communities, *Economica* 68, 587-606.

Miranda, A. and Zhu, Y. (2013) English deficiency and the native-immigrant wage gap, *Economics Letters* 118, 38-41.

Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MIT Press, Massachusetts.

Yao, Y. and van Ours, J.C. (2015) Language skills and labor market performance of immigrants in the Netherlands, *Labour Economics* 34, 76-85.

Yao, Y. and van Ours, J.C. (2019) Dialect speech and wage, *Economics Letters* 177, 35-38.

Table 1: Summary statistics

	Full sample	Wage sample
In employment	0.46	
Log real hourly wage		2.19
EAL	0.66	0.57
No qualification	0.24	0.13
Below GCSE/O-Level	0.08	0.08
GCSE/O-Level	0.14	0.12
A-Level	0.11	0.13
Higher Education Diploma	0.11	0.14
First Degree	0.16	0.20
Higher Degree	0.17	0.21
Highest qualification is foreign	0.28	0.32
Age	38.80	39.10
White	0.23	0.32
Mixed	0.02	0.02
Asian	0.40	0.30
Black	0.10	0.13
Other Ethnicity	0.26	0.23
Christian	0.37	0.51
Buddhist	0.02	0.02
Hindu	0.12	0.14
Muslim	0.35	0.16
Other Religion	0.06	0.07
No Religion	0.06	0.07
Religion missing	0.03	0.03
London	0.44	0.42
Southeast	0.10	0.13
Rest of England	0.41	0.39
Wales	0.01	0.01
Scotland	0.03	0.04
Northern Ireland	0.02	0.02
Observations	2,847	1,308

Table 2: Log-wage equations, wage sample (N=1,308)

	(1)	(2)	(3)	(4)	(5)
EAL	-0.192***	-0.207	-0.170***	-0.142***	-0.134***
	(0.033)	(0.032)	(0.033)	(0.034)	(0.028)
Age-at-arrival 10-15					-0.113*
					(0.067)
Age-at-arrival 16-29					-0.057
					(0.034)
Age-at-arrival 30+					-0.188
					(0.101)
Age, region and wave controls	No	Yes	Yes	Yes	Yes
Highest qualification dummies	No	No	Yes	Yes	Yes
Ethnicity & religion dummies	No	No	No	Yes	Yes
R^2	0.025	0.076	0.167	0.187	0.194

Note: Robust standard errors in parentheses; ***, ** and * denote significance at 1%, 5% and 10% level. Other controls include variables listed in footnote to Table 4.

Table 3: Biprobit of EAL and Selection into Employment Estimates, full sample (N=2,847)

	EAL	Employment
EAL		0.102
		(0.174)
Age-at-arrival 10-15	-1.622***	-0.006
	(0.225)	(0.109)
Age-at-arrival 16-29	-1.550***	-0.125*
	(0.194)	(0.067)
Age-at-arrival 30+	-1.672***	-0.074
	(0.215)	(0.101)
Exclusion restrictions:		
Born in non-English-speaking country * (age-at-arrival>9)	2.316***	
	(0.189)	
Labour Force Participation Rate Female-Male Ratio		0.395**
		(0.159)
Secondary Education Attainment Female-Male Ratio		-0.198
		(0.167)
ρ	-0.105	
(p-value)	(0.103)	

Note: Robust standard errors in parentheses; ***, ** and * denote significance at 1%, 5% and 10% level. Other controls include variables listed in footnote to Table 4.

Table 4: Wage estimates, full sample (N=2,847/1,308)

	OLS	2SLS	Wooldridge
EAL	-0.134***	-0.327***	-0.305***
	(0.034)	(0.086)	(0.106)
Age-at-arrival 10-15	-0.113*	-0.098	-0.102
	(0.067)	(0.067)	(0.074)
Age-at-arrival 16-29	-0.057	-0.03	-0.006
	(0.035)	(0.036)	(0.046)
Age-at-arrival 30+	0.188*	-0.167*	-0.154
	(0.101)	(0.099)	(0,108)
Inverse Mills Ratio (IMR)			-0.44
			(0.376)

Note: Standard errors for Wooldridge estimates bootstrapped with 1000 repetitions. Robust standard errors for OLS and 2SLS in parentheses; ***, ** and * denote significance at 1%, 5% and 10% level. Other controls include age, age squared, region dummies, wave dummies, highest qualification dummies, ethnicity dummies and religion dummies.